Claims

What is claimed is:

1.   A method used when summarizing one or more documents, the method
5 comprising the steps of:

determining topicality scores for a plurality of phrasal expressions in the
one or more documents;

determining specificities for the plurality of phrasal expressions; and

determining, by using the topicality scores and specificities, an order for
10 the plurality of phrasal expressions,

whereby the order may be used when summarizing the one or more
documents.

2.   The method of claim 1, wherein the one or more documents are from a
15 collection of documents.

3.   The method of claim 1, further comprising the step of determining
phrase-phrase relations between given pairs of the phrasal expressions, and wherein the
step of determining, by using the topicality scores and specificities, an order for the
20 plurality of phrasal expressions further comprises the step of determining, by using the
topicality scores, specificities and phrase-phrase relations, an order for the plurality of
phrasal expressions,

4.   The method of claim 1, wherein given ones of the plurality of phrasal
25 expressions comprise one or more of following: noun phrases, noun phrases with
corresponding prepositional phrases, subject-verb pairs and verb-object pairs.

5.   The method of claim 1, further comprising the step of determining the
phrasal expressions from the one or more documents.

YOR920030337US1   -14-

6.　　　　The method of claim 1, wherein the step of determining topicality scores further comprises the steps of:

determining one or more document vectors from the one or more

5　documents;

determining one or more phrase vectors from the one or more phrasal expressions;

using the one or more document vectors to determine a subspace;

using the subspace to determine one or more subspace-based document

10　vectors from the one or more document vectors;

using the subspace to determine one or more subspace-based phrase vectors from the one or more phrase vectors; and

computing topicality scores by using the one or more subspace-based document vectors and the one or more subspace-based phrase vectors.

15

7.　　　　The method of claim 1, wherein the step of determining specificities for the plurality of phrasal expressions further comprises the step of:

determining a specificity for each pair of phrasal expressions in the plurality of phrasal expressions.

20

8.　　　　The method of claim 7, wherein each specificity indicates an order between a pair of phrasal expressions in the plurality of phrasal expressions.

9.　　　　The method of claim 7, wherein the step of determining specificity for

25　each pair of phrasal expressions in the plurality of phrasal expressions further comprises the step of:

determining specificity for each pair of phrasal expressions in the plurality of phrasal expressions by using one or more of the following: set inclusion and an ontology.

10.   The method of claim 7, wherein each specificity, for a pair of phrasal expressions, indicates whether a first phrasal expression in the pair is more specific than a second phrasal expression in the pair, whether the first phrasal expression in the pair is less specific than the second phrasal expression in the pair, or whether specificity is undefined for the first and second phrasal expressions.

11.   The method of claim 7, wherein the step of determining specificity for each pair of phrasal expressions in the plurality of phrasal expressions further comprises the steps of:

        generating two content word sets from a given pair of phrasal expressions;

        using the content word sets to determine specificity corresponding to the given pair of phrasal expressions.

12.   The method of claim 7, wherein the specificity for a given pair of phrasal expressions indicates a specificity order between the phrasal expressions in the given pair.

13.   The method of claim 1, wherein the step of determining the order further comprises determining, by using the topicality scores and specificities, one or more parent-sibling relationships between the plurality of phrasal expressions.

14.   The method of claim 13, wherein the step of determining an order for the plurality of phrasal expressions further comprises the steps of:

        using the parent-sibling relationships to determine a phrasal expression tree; and

        displaying a representation of the phrasal expression tree.

15.   The method of claim 14, wherein the representation of the phrasal expression tree comprises a number of nodes, each node corresponding to a phrasal expression.

5   16.   The method of claim 15, wherein a portion of the nodes are parent nodes and connect to sibling nodes, and wherein the step of displaying a representation of the phrasal expression tree further comprises the steps of:

        expanding a parent node when the parent node is accessed and the parent node is contracted; and

10   contracted a parent node when the parent node is accessed and the parent node is expanded.

17.   The method of claim 13, wherein a specificity is defined or undefined for pairs of phrasal expressions, wherein the step of determining one or more parent-sibling
15 relationships further comprises the step of:

        selecting a pair of phrasal expressions having a defined specificity;

        determining if a selected one of the phrasal expressions in the selected pair has a parent assigned to the selected phrasal expression;

        assigning the other phrasal expression in the pair to be the parent of the
20 selected phrasal expression when the selected phrasal expression does not have a parent assigned to the selected phrasal expression;

        when the selected phrasal expression does have a parent assigned to the selected phrasal expression, determining whether the other of the phrasal expressions in the selected pair should be assigned as the parent of the selected phrasal expression based
25 on one or more of the following: content words for each of the phrasal expressions in the selected pair and topicality for each of the phrasal expressions in the selected pair.

18.   An apparatus used when summarizing one or more documents, the apparatus comprising:

one or more memories; and

one or more processors coupled to the one or more memories, the one or more processors configured:

to determine topicality scores for a plurality of phrasal expressions in the one or more documents;

to determine specificities for the plurality of phrasal expressions; and

to determine, by using the topicality scores and specificities, an order for the plurality of phrasal expressions,

whereby the order may be used when summarizing the one or more documents.

19.   The apparatus of claim 18, wherein the one or more documents are from a collection of documents.

20.   The apparatus of claim 18, wherein the one or more processors are further configured to determine phrase-phrase relations between given pairs of the phrasal expressions, and wherein the one or more processors are further configured, when determining an order for the plurality of phrasal expressions, to determine, by using the topicality scores, specificities and phrase-phrase relations, an order for the plurality of phrasal expressions,

21.   The apparatus of claim 18, wherein given ones of the plurality of phrasal expressions comprise one or more of following: noun phrases, noun phrases with corresponding prepositional phrases, subject-verb pairs and verb-object pairs.

22.   The apparatus of claim 18, wherein the one or more processors are further configured to determine the phrasal expressions from the one or more documents.

23.   The apparatus of claim 18, wherein the one or more processors are further configured, when determining topicality scores:

        to determine one or more document vectors from the one or more documents;

        to determine one or more phrase vectors from the one or more phrasal expressions;

        to use the one or more document vectors to determine a subspace;

        to use the subspace to determine one or more subspace-based document vectors from the one or more document vectors;

        to use the subspace to determine one or more subspace-based phrase vectors from the one or more phrase vectors; and

        to compute topicality scores by using the one or more subspace-based document vectors and the one or more subspace-based phrase vectors.

24.   The apparatus of claim 18, wherein the one or more processors are further configured, when determining the order, to determine, by using the topicality scores and specificities, one or more parent-sibling relationships between the plurality of phrasal expressions.

25.   An article of manufacture for use when summarizing one or more documents, the article of manufacture comprising:

        a computer readable medium containing one or more programs which when executed implement the steps of:

        determining topicality scores for a plurality of phrasal expressions in the one or more documents;

        determining specificities for the plurality of phrasal expressions; and

        determining, by using the topicality scores and specificities, an order for the plurality of phrasal expressions,

whereby the order may be used when summarizing the one or more documents.